 ISBN : 978 – 979 – 16353 – 9 – 4

S - 14

PEMODELAN REGRESI POISSON, BINOMIAL NEGATIF DAN PADA KASUS KECELAKAAN KENDARAAN BERMOTOR DI LALU LINTAS SUMATERA BARAT

Irwan[1], Devni Prima Sari[2]
[1,2] Jurusan Matematika FMIPA Univ. Negeri Padang
[1]irwan.math.165@gmail.com, [2]devniprimasari@yahoo.co.id

Abstrak

Tingginya angka kecelakaan kendaraan bermotor di lalu lintas harusnya menjadi topik pembicaraan yang harus disingkapi segera mungkin oleh segala pihak. Hal ini juga didukung oleh **fakta** mencengangkan yang dikabarkan Dinas Perhubungan Indonesia, yaitu kecelakaan lalu lintas menjadi penyebab kematian nomor tiga di Indonesia setelah serangan jantung dan strok. Salah satu cara untuk menyingkapi kasus tingginya jumlah kecelakaan kendaraan bermotor ini dilakukan dengan mengetahui faktor-faktor yang paling dominan yang menyebabkan terjadinya kecelakaan bermotor. Analisis faktor penyebab dominan dan memperkirakan jumlah kecelakaan bisa dilakukan dengan menggunakan model regresi Poisson. Regresi Poisson mempunyai asumsi *equi-dispersion*, yaitu kondisi dimana nilai rataan dan variansi dari variabel respon bernilai sama. Pada kenyataannya, pada data sering dijumpai variansi dari variabel respon lebih besar nilai rataannya (overdispersi). Untuk mengatasi permasalahan tersebut digunakan model regresi Binomial Negatif.

Kata kunci: angka kecelakaan, faktor dominan, regresi Poison, binomial negatif

A. PENDAHULUAN

Kecelakaan lalu lintas merupakan masalah yang umum terjadi dalam penyelenggaraan sistem transportasi di banyak negara. Pada negara-negara berkembang, termasuk di Indonesia, kecelakaan lalu lintas ini cenderung mengalami peningkatan. PT Jasa Raharja (persero) mencatat, setiap 15 menit terdapat satu orang yang meninggal dunia di Indonesia karena kecelakaan lalu lintas. Ini terlihat dari tingginya biaya santunan kecelakaan yang diklaim kepada asuransi Jasa Raharja mencapai Rp 1 triliun setiap tahunnya. PT Jamsostek (Persero) mencatat klaim kecelakaan lalu lintas mendominasi dibanding kecelakaan kerja lainnya. Kepala PT Jamsostek (Persero) Cabang Tanjung Priok Muhammad Akip mengungkapkan, kecelakaan lalu lintas masih mendominasi daftar panjang kecelakaan kerja yang diklaim ke Jamsostek setiap tahunnya.

Jumlah kecelakaan yang tinggi akan menjadi salah satu faktor yang tidak menguntungkan bagi perusahaan asuransi kendaraan bermotor. Disamping itu, seluruh pihak harus lebih mewaspadai faktor-faktor yang menyebabkan terjadinya kecelakaan lalu lintas. Masih tingginya angka kecelakaan di jalan raya juga menjadi pekerjaan rumah besar bagi pemerintah.

Perang opini tentang faktor penyebab utama meningkatnya jumlah kecelakaan pun terjadi di kalangan masyarakat, sehingga perlu dilakukan perbaikan pada faktor-faktor yang berkontribusi dalam kecelakaan. Dalam hal ini faktor manusia memiliki kontribusi terbesar pada kecelakaan kendaraan bermotor, sehingga faktor tersebut sangat penting untuk diamati dalam upaya mengurangi terjadinya kecelakaan lalu lintas yang melibatkan kendaraan bermotor. Perbedaan karakteristik sosio-ekonomi, karakteristik pergerakan dan perilaku pengemudi

 ISBN : 978 – 979 – 16353 – 9 – 4

kendaraan bermotor menjadi dasar pertimbangan dalam identifikasi faktor-faktor penyebab terjadinya kecelakaan.

Model regresi Poisson dapat diterapkan untuk mengkaji sejauh mana faktor-faktor penyebab terjadinya kecelakaan berpengaruh terhadap jumlah terjadinya kecelakaan, serta sejauh mana dampak faktor tersebut terhadap jumlah kecelakaan bermotor di lalu lintas. Model regresi Poisson adalah suatu metode statistika yang digunakan untuk menganalisis hubungan antara variabel respon yang dapat dihitung (data cacah/count) dengan satu atau lebih variabel penjelas, dimana rataan dan variansinya sama. Pada prakteknya seringkali data cacah memperlihatkan variansi yang sangat besar, dimana variansi sampel lebih besar dari rataan sampel (overdispersion). Ketika model Poisson diaplikasikan untuk data overdispersi, menyebabkan standar error underestimate. Akhibatnya, beberapa variabel penjelas menjadi tidak signifikan. Oleh karena itu, sasaran dari penelitian ini untuk menggunakan model regresi Binomial Negatif sebagai solusi alternatif jika terjadi kasus overdispersi. Selanjutnya, model-model regresi Poisson, Binomial Negatif diuji dan dibandingkan pada jenis data jumlah kecelakaan kendaraan bermotor di Sumatera Barat.

B. METODE PENELITIAN

Penelitian ini merupakan penelitian terapan yang berkaitan dengan model regresi Poisson dan Binomial Negatif pada kasus kecelakaan kendaraan bermotor di lalu lintas pada propinsi Sumatera Barat. Setelah ditemukan model, maka dilakukan simulasi menggunakan data kecelakaan lalu lintas dengan factor penyebab utama adalah manusia sedangkan *rating factor* Kepemilikan SIM, Usia pengendara, Jenis kelamin pengendara, Pendidikan pengendara dan Kondisi fisik pengendara. Sementara itu sebagai *rating classes* adalah bagiian-bagian dari *rating factor*. Untuk simulasi model digunakan data kecelakaan lalintas di Sumatera Barat yang terdapat pada Direktorat Lalu Lintas Kepolisian Daerah Sumatera Barat. Data kecelakaan yang diambil adalah data tahun 2012.

C. HASIL DAN PEMBAHASAN

Model Regresi Poisson

Overdispersi

Overdispersi merupakan salah satu masalah yang sering terjadi dalam analisis regresi Poisson. Distribusi Poisson sering dianjurkan dalam menghitung data tetapi distribusi ini tidaklah mencukupi karena data menampilkan variansi yang lebih besar dari yang diprediksikan oleh Poisson. Hal ini diistilahkan dengan overdispersi atau variansi ekstra-Poisson. Overdispersi bisa terjadi karena pengelompokkan di dalam populasi dan pengukuran atau percobaan secara berulang pada objek yang sama (McCullagh & Nelder, 1989).

Ada atau tidaknya overdispersi dapat dilihat dari nilai *Deviance* atau *Pearson Chi-square* yang dibagi dengan derajat bebasnya. Apabila nilai *pearson Chi-square* dibagi dengan derajat bebas lebih besar daripada 1, ini menunjukkan nilai variansi yang lebih besar daripada rataan; overdispersi telah terjadi.

Permasalahan overdispersi biasanya terjadi pada kasus-kasus nyata. Untuk mengatasinya dapat dilakukan dua metode, yaitu:

1. Dengan mengansumsikan $Var(y_i) = \sigma^2\lambda_i$ dan mengestimasi parameter σ^2, yang kemudian disebut dengan model Quasi-*likelihood*.
2. Dengan mengubah distribusi variabel respon menjadi binomial negatif, dimana lebih terdispersi daripada Poisson, yang kemudian disebut dengan model regresi binomial negatif.

Distribusi Poisson

Misalkan Y_i merupakan variabel random hitungan klaim dalam i kelas, i = 1,2 ..... n, dimana n menunjukkan jumlah *rating classes*. Jika Y_i mengikuti distribusi Poisson, fungsi kepadatan peluang adalah,

 ISBN : 978 – 979 – 16353 – 9 – 4

$$Pr(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, y_i = 0,1, \ldots \tag{5.1}$$

dengan rataan dan variansi, $E(Y_i) = Var(Y_i) = \lambda_i$.

Bukti:

Berdasarkan deret Maclaurin dari e^u,

$$e^u = \sum_{y_i=0}^{\infty} \frac{u^{y_i}}{y_i!} = 1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots$$

maka:

$$\sum_{y_i=0}^{\infty} P(y_i) = e^{-\lambda_i} \sum_{y_i=0}^{\infty} \frac{\lambda_i^{y_i}}{y_i!} = e^{-\lambda_i}. e^{\lambda_i} = e^{\lambda_i-\lambda_i} = e^0 = 1$$

$$E(Y_i) = \sum_{y_i=0}^{\infty} y_i. P(Y_i) = \sum_{y_i=0}^{\infty} \frac{y_i e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} = \sum_{y_i=1}^{\infty} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{(y_i-1)!} = e^{-\lambda_i} \sum_{y_i=1}^{\infty} \frac{\lambda_i^{y_i}}{(y_i-1)!}$$

$$= \lambda_i. e^{-\lambda_i} \sum_{y_i=1}^{\infty} \frac{\lambda_i^{y_i-1}}{(y_i-1)!}.$$

Misalkan:

$$x_i = y_i - 1 \Rightarrow y_i = x_i + 1$$

maka:

$$E(Y_i) = \lambda_i. e^{-\lambda_i} \sum_{x_i=0}^{\infty} \frac{\lambda_i^{x_i}}{x_i!} = \lambda_i. e^{-\lambda_i}. e^{\lambda_i} = \lambda_i. e^0 = \lambda_i$$

$$Var(Y_i) = E(Y_i^2) - \left(E(Y_i)\right)^2$$

$$E(Y_i^2) = E\left(Y_i(Y_i-1)\right) + E(Y_i)$$

$$E\left(Y_i(Y_i-1)\right) = \sum_{y_i=0}^{\infty} y_i(y_i-1). P(Y_i)$$

$$= e^{-\lambda_i} \sum_{y_i=2}^{\infty} \frac{\lambda_i^{y_i}}{(y_i-2)!}$$

$$= \lambda_i^2. e^{-\lambda_i} \sum_{y_i=1}^{\infty} \frac{\lambda_i^{y_i-2}}{(y_i-2)!}.$$

Misalkan:

$$x_i = y_i - 2 \Rightarrow y_i = x_i + 2$$

maka:

$$E\left(Y_i(Y_i-1)\right) = \lambda_i^2. e^{-\lambda_i} \sum_{x_i=0}^{\infty} \frac{\lambda_i^{x_i}}{x_i!} = \lambda_i^2. e^{-\lambda_i}. e^{\lambda_i} = \lambda_i^2. e^0 = \lambda_i^2.$$

Jadi,

$$E(Y_i^2) = E\left(Y_i(Y_i-1)\right) + E(Y_i)$$

$$= \lambda_i^2 + \lambda_i$$

sehingga:

$$Var(Y_i) = E(Y_i^2) - \left(E(Y_i)\right)^2$$

$$= \lambda_i^2 + \lambda_i - (\lambda_i)^2$$

$$= \lambda_i.$$

Terlihat bahwa: $Y_i \sim Poisson(\lambda_i), E(Y_i) = Var(Y_i) = \lambda_i$ ■

Model Regresi Poisson

Dalam berbagai eksperimen, seringkali data cacah yang merupakan objek penelitian dipengaruhi oleh sejumlah variabel penjelas (*explanatory*). Sehingga untuk mengetahui pola hubungan kedua variabel tersebut, dapat digunakan suatu model regresi yang didasarkan pada distribusi Poisson. Pada regresi Poisson diasumsikan bahwa variabel dependen Y_i yang menyatakan jumlah (cacah) kejadian berdistribusi Poisson, diberikan sejumlah variabel independen $x_1, \dots, x_k$. Y_i mengikuti distribusi Poisson, fungsi kepadatan peluang adalah,

$$Pr(Y_i|x_1, \dots, x_k) = \frac{e^{-\lambda_i}{\lambda_i}^{y_i}}{y_i!}, y_i = 0,1, \dots$$

atau $Y_i \sim Poi(\lambda_i), i = 1,2,3, \dots, n$.

Salah satu tujuan dari analisis regresi adalah untuk menentukan pola hubungan antara variabel respon dengan variabel penjelas. Selanjutnya, dalam regresi Poisson hubungan tersebut dapat dituliskan dalam bentuk:

$$E[Y_i|x_i] = \lambda_i = \beta_0 + x_1\beta_1 + \dots + x_k\beta_k$$

atau dalam bentuk vektor ditulis sebagai

$$E[Y_i|x_i] = \lambda_i = \boldsymbol{x_i^T\beta} \quad \textbf{(5.2)}$$

Karena nilai $\lambda_i > 0$, maka digunakan fungsi link $\eta_i = \exp(\boldsymbol{x_i^T\beta})$ atau $\eta_i = log\lambda_i = \boldsymbol{x_i^T\beta}$ untuk menghubungkan $\lambda_i = E[Y_i|\boldsymbol{x_i}]$ dengan fungsi linear $\boldsymbol{x_i^T\beta}$, sehingga hubungan antara $\lambda_i = E[Y_i|\boldsymbol{x_i}]$ dan $\boldsymbol{x_i^T\beta}$ menjadi tepat. Dengan demikian, model regresi dapat ditulis dalam bentuk:

$$E[Y_i|\boldsymbol{x_i}] = \lambda_i = exp(\boldsymbol{x_i^T\beta}), i = 1,2, \dots, n. \quad (5.3)$$

Dengan $\boldsymbol{\beta}$ merupakan parameter yang tidak diketahui dalam model dan harus diestimasi. Dimana $\boldsymbol{x_i}$ merupakan vektor *p* x 1 dari variabel penjelas, dan $\boldsymbol{\beta}$ merupakan vektor *p* x 1 dari parameter regresi.

Estimasi Parameter

Untuk mengestimasi parameter-parameter dalam regresi Poisson dapat digunakan metode estimasi maksimum *likelihood* (MLE). Metode estimasi maksimum *likelihood* dapat dilakukan jika distribusi data diketahui.

Langkah pertama yang dilakukan adalah menentukan fungsi *likelihood* dari model regresi Poisson. Dengan mengasumsikan $Y_1, Y_2, \dots, Y_n$ adalah variabel random yang *mutually independent* atau Y_i ~ Poisson (λ_i), maka fungsi *likelihood* untuk model regresi Poisson adalah sebagai berikut:

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} P(y_i;\boldsymbol{\beta}) \\ &= \prod_{i=1}^{n} \frac{e^{-\lambda_i}{\lambda_i}^{y_i}}{y_i!} \\ &= \frac{exp[-\sum_{i=1}^{n}\lambda_i][\prod_{i=1}^{n}{\lambda_i}^{y_i}]}{\prod_{i=1}^{n} y_i!}. \end{aligned}$$

Selanjutnya dari fungsi *likelihood* diambil nilai lognya sehingga diperoleh fungsi *log-likelihood* dari persamaan di atas sebagai berikut:

$$\begin{aligned} log\, L(\boldsymbol{\beta}) &= log\left\{\prod_{i=1}^{n} P(y_i;\boldsymbol{\beta})\right\} \\ &= log\left\{\frac{exp[-\sum_{i=1}^{n}\lambda_i][\prod_{i=1}^{n}{\lambda_i}^{y_i}]}{\prod_{i=1}^{n} y_i!}\right\} \\ &= -\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n} y_i log\lambda_i - \sum_{i=1}^{n} log y_i!\,. \end{aligned} \quad (5.4)$$

Dari persamaan (5.2)) bahwa $\lambda_i = exp(\boldsymbol{x_i}^T\boldsymbol{\beta})$ dengan $\boldsymbol{x_i}$ adalah nilai-nilai kovariat untuk observasi ke-i, maka diperoleh persamaan sebagai berikut:

$$log\, L(\boldsymbol{\beta}) = -\sum_{i=1}^{n} exp(\boldsymbol{x_i}^T\boldsymbol{\beta}) + \sum_{i=1}^{n} y_i log[exp(\boldsymbol{x_i}^T\boldsymbol{\beta})] - \sum_{i=1}^{n} log y_i!\,.$$

(5.5)

Kemudian persamaan (5.5) diturunkan terhadap β_j dan disamakan dengan nol, yaitu:

$$\frac{\partial\, \mathrm{l}(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} y_i x_{ij} \frac{exp(\boldsymbol{x_i}^T\boldsymbol{\beta})}{exp(\boldsymbol{x_i}^T\boldsymbol{\beta})} - \sum_{i=1}^{n} exp(\boldsymbol{x_i}^T\boldsymbol{\beta}) x_{ij} = 0$$

$$= \sum_{i=1}^{n} y_i x_{ij} - \sum_{i=1}^{n} \lambda_i x_{ij} = 0$$

$$\frac{\partial\, \mathrm{l}(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n}(y_i - \lambda_i)x_{ij} = 0 \;, j = 1,2, \dots, p. \qquad (5.6)$$

Setelah disamakan dengan nol, maka akan terdapat p (sejumlah parameter yang ada) persamaan. Dalam persamaan (5.6) terdapat suku $exp(\boldsymbol{X_i}^T\boldsymbol{\beta})$ sehingga bentuk pasti (*closed form*) dari β sulit ditentukan. Oleh karena itu, untuk mengestimasi parameter β dilakukan secara iteratif dengan bantuan komputer yang didasarkan pada suatu prosedur (algoritma) iterasi yang disebut dengan *Iteratively Weighted Least Square* (IWLS).

Untuk mempermudah penerapan pada regresi IWLS Poisson, notasi rataan Poisson, λ_i, akan diganti dengan μ_i . Dalam rangka mencari $\hat{\beta}$ pada regresi Poisson log $L(\beta)$ dapat dimaksimalkan dengan menggunakan metode WLS.

$$\frac{\partial\, \mathrm{l}(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i} \frac{y_i - \mu_i}{Var(Y_i)} \frac{\partial \mu_i}{\partial \boldsymbol{\beta_j}} = \sum_{i} \frac{(y_i - \mu_i)}{\mu_i} x_{ij}\mu_i$$

$$\sum_{i}(y_i - \mu_i)x_{ij} = 0 \quad j = 1,2, \dots, p, \qquad (5.7)$$

menghasilkan perkiraan kuadrat terkecil, $\tilde{\boldsymbol{\beta}}$.
Hal ini menunjukkan bahwa hasil estimasi dengan MLE dan WLS sama.

Uji Ketepatan (*Goodness of Fit*) Model Regresi Poisson

Uji ketepatan model regresi Poisson dilakukan dengan dua cara yaitu, *Pearson Chi-square statistic* dan *Deviance*.

Pearson Chi-square statistic untuk Model Regresi Poisson

Ukuran lain yang bisa digunakan untuk uji *goodness of fit* yaitu *Pearson Chi-square statistic* (McCullagh & Nelder, 1989) yang didefinisikan sebagai

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{Var(\mu_i)}$$

dimana $Var(\mu_i)$ yaitu fungsi variansi dari distribusi peluang datanya dan $Var(\mu_i) = \frac{Var(Y_i)}{\phi}$.
Karena data yang diobservasi itu berdistribusi Poisson, maka $Var(\mu_i) = \mu_i$ sehingga *Statistik Pearson Chi-square* untuk regresi Poissonnya yaitu:

 ISBN : 978 – 979 – 16353 – 9 – 4

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i}.$$

(5.8)

Deviance untuk Model Regresi Poisson

Deviance dapat diartikan sebagai logaritma dari uji *likelihood*nya, yaitu:

$$D = -2log\left[\frac{L(\boldsymbol{y};\boldsymbol{\mu})}{L(\boldsymbol{y};\boldsymbol{y})}\right] = 2(logL(\boldsymbol{y};\boldsymbol{y}) - logL(\boldsymbol{y};\boldsymbol{\mu})).$$

(5.9)

Dalam persamaan tersebut, $L(\boldsymbol{y};\boldsymbol{\mu})$ adalah fungsi *likelihood current model* sedangkan $L(\boldsymbol{y};\boldsymbol{y})$ adalah fungsi *likelihood* saturated model dari distribusi Poisson. Adapun fungsi log-*likelihood current model* dituliskan sebagai berikut:

$$\begin{aligned} l(\boldsymbol{\mu},\boldsymbol{y}) &= logL(\boldsymbol{\mu},\boldsymbol{y}) \\ &= log\prod_{i=1}^{n}\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \\ &= \sum_{i=1}^{n}(y_i log\mu_i - \mu_i - logy_i!). \end{aligned}$$

(5.10)

Sedangkan untuk *saturated model*, dimana nilai-nilai μ_i diganti dengan nilai y_i (tanpa asumsi tentang keeratan hubungannya dengan variabel *x* nya), fungsi *likelihood saturated model*-nya yaitu:

$$\begin{aligned} L(\boldsymbol{y};\boldsymbol{y}) &= \prod_{i=1}^{n} f(y_i, y_i) \\ &= \prod_{i=1}^{n}\frac{e^{-y_i}y_i^{y_i}}{y_i!}. \end{aligned}$$

Sehingga fungsi log-*likelihood saturated model*-nya menjadi

$$\begin{aligned} l(\boldsymbol{y};\boldsymbol{y}) &= logL(\boldsymbol{y};\boldsymbol{y}) \\ &= log\prod_{i=1}^{n}\frac{e^{-y_i}y_i^{y_i}}{y_i!} \\ &= \sum_{i=1}^{n}(y_i logy_i - y_i - logy_i!). \end{aligned}$$

(5.11)

Sehingga *Deviance*, D, bisa diperoleh dengan mensubstitus (5.10) dan (5.11) ke dalam (5.9) dan diperoleh

$$\begin{aligned} D &= 2\left\{\sum_{i=1}^{n}(y_i logy_i - y_i - logy_i!) - \sum_{i}^{\mathrm{n}}(y_i log\mu_i - \mu_i - logy_i!)\right\} \\ &= 2\sum_{i=1}^{n}\left\{y_i \log\frac{y_i}{\mathrm{\mu_i}} - (\mathrm{y_i} - \mathrm{\mu_i})\right\}. \end{aligned}$$

(5.12)

 ISBN : 978 – 979 – 16353 – 9 – 4

Karena $\sum_{i=1}^{n}(y_i - \mu_i) = 0$, maka persamaan (5.12) menjadi:

$$D = 2\sum_{i=1}^{n}\left\{y_i log\frac{y_i}{\mu_i}\right\}.$$

(5.13)

Model Regresi Binomial Negatif

Overdispersi Metode Binomial Negatif

Berdasarkan Poisson, rataan, λ_i, diasumsikan konstan atau homogen dalam kelas. Dengan asumsi λ_i untuk Gamma dengan rataan $E(\lambda_i) = \mu_i$ dan varians $Var(\lambda_i) = \mu_i{}^2 v_i{}^{-1}$, dan $Y_i|\lambda_i$ menjadi Poisson dengan rataan bersyarat $E(Y_i|\lambda_i) = \lambda_i$ dapat ditunjukkan bahwa distribusi marjinal Y_i mengikuti distribusi binomial negatif dengan fungsi kepadatan peluang,

$$Pr(Y_i = y_i) = \int Pr(Y_i = y_i|\lambda_i)f(\lambda_i)d\lambda_i$$
$$= \frac{\Gamma(y_i + v_i)}{\Gamma(y_i + 1)\Gamma(v_i)}\left(\frac{v_i}{v_i + \mu_i}\right)^{v_i}\left(\frac{\mu_i}{v_i + \mu_i}\right)^{y_i}$$

(5.14)

di mana rataan $E(Y_i) = \mu_i$, dan variansi adalah $Var(Y_i) = \mu_i + \mu_i{}^2 v_i{}^{-1}$.

Apabila overdispersi terjadi pada data yang dimodelkan dengan regresi Poisson, maka salah satu jalan yang dapat diambil adalah memodelkan ulang data tersebut dengan model yang lebih terdispersi. Dalam hal ini, model yang digunakan adalah model binomial negatif.

Misalkan $Y_i \sim$ Poisson(λ_i), terhadap λ_i itu sendiri adalah variabel random dengan distribusi Gamma. Misalkan

$Y_i|\lambda_i \sim$ Poisson(λ_i)

$\lambda_i \sim$ Gamma $\left(v_i, \frac{\mu_i}{v_i}\right)$

dengan Gamma $\left(v_i, \frac{\mu_i}{v_i}\right)$ adalah distribusi Gamma dengan rataan $E(\lambda_i) = \mu_i$ dan variansi

$Var(\lambda_i) = \mu_i{}^2 v_i{}^{-1}$, dengan fungsi densitas $Pr(\lambda_i) = \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i}\Gamma(v_i)}\lambda_i{}^{v_i-1}exp\left(\frac{-\lambda_i}{\left(\frac{\mu_i}{v_i}\right)}\right)$ untuk $\lambda_i > 0$ dan nol untuk yang lain.

Dapat ditunjukkan bahwa distribusi tak bersyarat dari y_i adalah binomial negatif, sebagai berikut:

Fungsi kepadatan peluang bersyarat dari y_i adalah $Pr(Y_i = y_i|\lambda_i) = \frac{e^{-\lambda_i}\lambda_i{}^{y_i}}{y_i!}$.

Fungsi kepadatan peluang dari λ_i adalah

$$Pr(\lambda_i) = \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i}\Gamma(v_i)}\lambda_i{}^{v_i-1}exp\left(\frac{-\lambda_i v_i}{\mu_i}\right).$$

Memanfaatkan definisi dari fungsi densitas bersyarat, maka didapat fungsi peluang bersama dari y_i dan λ_i adalah

$$Pr(y_i, \lambda_i) = Pr(\lambda_i)Pr(Y_i = y_i|\lambda_i)$$
$$= \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i}\Gamma(v_i)}\lambda_i{}^{v_i-1}e^{\left(\frac{-\lambda_i v_i}{\mu_i}\right)}.\frac{e^{-\lambda_i}\lambda_i{}^{y_i}}{y_i!}$$

ISBN : 978 – 979 – 16353 – 9 – 4

$$= \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i) y_i!} \lambda_i^{v_i+y_i-1} e^{-\left(1+\frac{v_i}{\mu_i}\right)\lambda_i}.$$

Dengan diperolehnya fungsi peluang bersama dari y_i dan λ_i, maka fungsi peluang tak bersyarat dari y_i adalah

$$Pr(Y_i = y_i) = \int_0^\infty \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i) y_i!} \lambda_i^{v_i+y_i-1} e^{-\left(1+\frac{v_i}{\mu_i}\right)\lambda_i} \, d\lambda_i$$

$$= \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i) y_i!} \int_0^\infty \lambda_i^{v_i+y_i-1} e^{-\left(1+\frac{v_i}{\mu_i}\right)\lambda_i} \, d\lambda_i.$$

Misalkan $z_i = \left(1 + \frac{v_i}{\mu_i}\right)\lambda_i$ maka $dz_i = \left(1 + \frac{v_i}{\mu_i}\right) d\lambda_i$ dan $\lambda_i = \left(\frac{\mu_i}{v_i+\mu_i}\right) z_i$, maka dengan demikian didapatkan

$$Pr(Y_i = y_i) = \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i) y_i!} \int_0^\infty \left[\left(\frac{\mu_i}{v_i + \mu_i}\right) z_i\right]^{v_i+y_i-1} \left(\frac{\mu_i}{v_i + \mu_i}\right) e^{-z_i} \, dz_i$$

$$= \frac{\Gamma(y_i+v_i)}{\Gamma(y_i + 1)\Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i}\right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i}\right)^{y_i}.$$

Untuk $y_i = 0, 1, 2, \ldots$, $P(Y_i = y_i)$ adalah fungsi kepadatan peluang binomial negatif.

Harga ekspektasi/rataan dari distribusi ini adalah

$$E(Y_i) = E[E(Y_i|\lambda_i)] = E(\lambda_i) = v_i . \frac{\mu_i}{v_i} = \mu_i$$

$$E[Var(Y_i|\lambda_i)] = E\{E(Y_i^2|\lambda_i) - [E(Y_i|\lambda_i)]^2\}$$

$$E[\lambda_i] = \{E(Y_i^2) - E[E(Y_i|\lambda_i)]^2\}$$

$$E[\lambda_i] = E(Y_i^2) - E(\lambda_i^2)$$

$$E(Y_i^2) = E[\lambda_i] + E(\lambda_i^2)$$

$$= v_i . \frac{\mu_i}{v_i} + \left[v_i^2 \left(\frac{\mu_i}{v_i}\right)^2 + v_i \left(\frac{\mu_i}{v_i}\right)^2\right]$$

$$= \mu_i + \mu_i^2 + \frac{\mu_i^2}{v_i}.$$

Dengan variansinya adalah

$$Var(Y_i) = E(Y_i^2) - [E(Y_i)]^2$$

$$= \left(\mu_i + \mu_i^2 + \frac{\mu_i^2}{v_i}\right) - \mu_i^2$$

$$= \mu_i + \mu_i^2 v_i^{-1}.$$

Distribusi Binomial Negatif

Parameter berbeda dapat menghasilkan berbagai jenis distribusi binomial negatif. Misalnya, dengan mengambil $v_i = a^{-1}$, Y_i mengikuti sebuah distribusi binomial negatif dengan rataan $E(Y_i) = \mu_i$ dan variansi $Var(Y_i) = \mu_i(1 + a\mu_i)$, di mana a menunjukkan parameter dispersi (Lawless, 1987); (Cameron & Trivedi, 1986). Sehingga persamaan (5.14) menjadi,

$$Pr(Y_i = y_i) = \int Pr(Y_i = y_i|\lambda_i) f(\lambda_i) d\lambda_i$$

 ISBN : 978 – 979 – 16353 – 9 – 4

$$= \frac{\Gamma(y_i + a^{-1})}{\Gamma(y_i + 1)\Gamma(a^{-1})}\left(\frac{a^{-1}}{a^{-1} + \mu_i}\right)^{a^{-1}}\left(\frac{\mu_i}{a^{-1} + \mu_i}\right)^{y_i}$$

$$Pr(Y_i = y_i) = \frac{\Gamma(y_i + a^{-1})}{y_i!\,\Gamma(a^{-1})}\left(\frac{a\mu_i}{1 + a\mu_i}\right)^{y_i}(1 + a\mu_i)^{-a^{-1}}. \tag{5.15}$$

Jika a sama dengan nol, rataan dan variansi akan sama, E(Y_i) = Var(Y_i), akan menjadi distribusi Poisson. Jika a > *0*, variansi akan melebihi rataan, Var(Y_i) > E(Y_i), dan distribusi memungkinkan overdispersi. Dalam tulisan ini, distribusi akan disebut sebagai binomial negatif.

Estimasi parameter untuk model regresi binomial negatif

Untuk mengestimasi parameter β dan a dalam regresi binomial negatif dapat digunakan metode estimasi maksimum *likelihood* (MLE). Langkah pertama yang dilakukan adalah menentukan fungsi *likelihood* dari model regresi binomial negatif. Jika diasumsikan bahwa rataan atau *fitted value* adalah multiplikatif, yaitu, $E(Y_i|\boldsymbol{x_i}) = \mu_i = e_i exp(\boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta})$, maka fungsi *likelihood* untuk model regresi binomial negatif adalah sebagai berikut:

$$L(\boldsymbol{\beta}, a) = \prod_{i=1}^{n} P(\boldsymbol{\beta}, a)$$

$$= \prod_{i=1}^{n} \frac{\Gamma(y_i + a^{-1})}{y_i!\,\Gamma(a^{-1})}\left(\frac{a\mu_i}{1 + a\mu_i}\right)^{y_i}(1 + a\mu_i)^{-a^{-1}}$$

$$= \prod_{i=1}^{n} \frac{\Gamma(y_i + a^{-1})}{y_i!\,\Gamma(a^{-1})}\prod_{i=1}^{n}\left(\frac{a\mu_i}{1 + a\mu_i}\right)^{y_i}\prod_{i=1}^{n}(1 + a\mu_i)^{-a^{-1}}.$$

Selanjutnya dari fungsi *likelihood* diambil nilai lognya sehingga diperoleh fungsi log-*likelihood* dari persamaan di atas sebagai berikut:

$$log\, L(\boldsymbol{\beta}, a) = log\left\{\prod_{i=1}^{n} P(\boldsymbol{\beta}, a)\right\}$$

$$= log\left\{\prod_{i=1}^{n} \frac{\Gamma(y_i + a^{-1})}{y_i!\,\Gamma(a^{-1})}\prod_{i=1}^{n}\left(\frac{a\mu_i}{1 + a\mu_i}\right)^{y_i}\prod_{i=1}^{n}(1 + a\mu_i)^{-a^{-1}}\right\}$$

$$= \sum_{i=1}^{n} log\left[\frac{\Gamma(y_i + a^{-1})}{\Gamma(a^{-1})}\right] - \sum_{i=1}^{n} log(y_i!) + \sum_{i=1}^{n} y_i\, log(a\mu_i) - \sum_{i=1}^{n} y_i\, log(1 + a\mu_i) - \sum_{i=1}^{n} a^{-1} log(1 + a\mu_i)$$

$$log\, L(\boldsymbol{\beta}, a) = \sum_{i=1}^{n} log\left[\frac{\Gamma(y_i + a^{-1})}{\Gamma(a^{-1})}\right] - \sum_{i=1}^{n} log(y_i!) + \sum_{i=1}^{n} y_i\, log(a\mu_i) - \sum_{i=1}^{n}(y_i + a^{-1})\, log(1 + a\mu_i).$$

Menurut Lawless (1987), $\frac{\Gamma(I+k)}{\Gamma(k)} = k \times (1 + k) \times ... \times (I - 1 + k)$ untuk I bilangan bulat. Sehingga, $\frac{\Gamma(y_i+a^{-1})}{\Gamma(a^{-1})} = a^{-1} \times (1 + a^{-1}) \times ... \times (y_i - 1 + a^{-1})$.

Maka, $log\, L(\boldsymbol{\beta}, a)$ bisa ditulis tanpa fungsi Gamma dengan

$$log\left[\frac{\Gamma(y_i + a^{-1})}{\Gamma(a^{-1})}\right] = log a^{-1} + \log(1 + a^{-1}) + \cdots + log(y_i - 1 + a^{-1})$$

$$log\left[\frac{\Gamma(y_i + a^{-1})}{\Gamma(a^{-1})}\right] = \sum_{r=0}^{y_i-1} log\left(\frac{1+ar}{a}\right).$$

Likelihood untuk model regresi binomial negatif I dapat ditulis sebagai,

$$log\ L(\boldsymbol{\beta}, a) = \sum_{i=1}^{n}\left\{\left[\sum_{r=0}^{y_i-1} log\left(\frac{1+ar}{a}\right)\right] - log(y_i!) + y_i\ log(a\mu_i) - (y_i + a^{-1})\ log(1 + a\mu_i)\right\}$$

$$l(\boldsymbol{\beta}, a) = \sum_{i=1}^{n}\left\{\left[log\sum_{r=0}^{y_i-1}(1+ar)\right] - y_i \log(a) - log(y_i!) + y_i\ log(a\mu_i) - (y_i + a^{-1})\ log(1 + a\mu_i)\right\}. \tag{5.16}$$

Oleh karena itu, estimasi maksimum *likelihood*, $(\widehat{\boldsymbol{\beta}}, \hat{a})$, dapat diperoleh dengan memaksimalkan $l(\boldsymbol{\beta}, a)$terhadap $\boldsymbol{\beta}$ dan a. Persamaan terkait adalah,

$$\begin{aligned}\frac{\partial l(\boldsymbol{\beta}, a)}{\partial \beta_j} &= \sum_{i=1}^{n}\left\{\frac{y_i a x_{ij}\mu_i}{a\mu_i} - \frac{(y_i + a^{-1})a x_{ij}\mu_i}{1 + a\mu_i}\right\}\\ &= \sum_{i=1}^{n}\left\{y_i - \frac{(a y_i\mu_i + \mu_i)}{1 + a\mu_i}\right\}x_{ij}\\ &= \sum_{i=1}^{n}\left\{\frac{y_i + a y_i\mu_i - a y_i\mu_i - \mu_i}{1 + a\mu_i}\right\}x_{ij}\\ &= \sum_{i=1}^{n}\frac{(y_i - \mu_i)x_{ij}}{1 + a\mu_i} = 0, \quad j = 1,2,\ldots,p,\end{aligned} \tag{5.17}$$

dan,

$$\frac{\partial l(\boldsymbol{\beta}, a)}{\partial a} = \sum_{i=1}^{n}\left\{\left[\sum_{r=1}^{y_i-1}\left(\frac{r}{1+ar}\right)\right] - \frac{y_i}{a} + \frac{y_i\mu_i}{a\mu_i}\left[-\frac{(y_i + a^{-1})\mu_i}{(1 + a\mu_i)} + (-a^{-2})log(1 + a\mu_i)\right]\right\}$$

$$\frac{\partial l(\boldsymbol{\beta}, a)}{\partial a} = \sum_{i=1}^{n}\left\{\left[\sum_{r=1}^{y_i-1}\left(\frac{r}{1+ar}\right)\right] + a^{-2}\log(1 + a\mu_i) - \frac{(y_i + a^{-1})\mu_i}{(1 + a\mu_i)}\right\} = 0. \tag{5.18}$$

Estimasi maksimum *likelihood*, $(\widehat{\boldsymbol{\beta}}, \hat{a})$, dapat diselesaikan secara bersamaan, dan melibatkan prosedur iterasi yang berurutan. Urutan pertama, dengan menggunakan nilai awal dari a, yaitu $a_{(0)}$, $l(\boldsymbol{\beta}, a)$ akan maksimal terhadap $\boldsymbol{\beta}$, menghasilkan $\boldsymbol{\beta}_{(1)}$. Persamaan (5.17) akan setara dengan kuadrat terbobot. Oleh karena itu, dengan sedikit modifikasi, permasalahan ini dapat dilakukan dengan menggunakan regresi IRWLS mirip dengan Poisson. Pada urutan kedua, dengan $\boldsymbol{\beta}$ tetap pada $\boldsymbol{\beta}_{(1)}$, $l(\boldsymbol{\beta}, a)$ adalah dimaksimalkan terhadap a, menghasilkan $a_{(1)}$. Persamaan terkait adalah persamaan (5.18), dan permasalahan dapat dilakukan dengan menggunakan iterasi Newton-Raphson. Iterasi dengan a dan $\boldsymbol{\beta}$ tetap, MLE, $(\widehat{\boldsymbol{\beta}}, \hat{a})$, akan diperoleh.

ISBN : 978 – 979 – 16353 – 9 – 4

Pendekatan yang lebih mudah untuk mengestimasi adalah dengan menggunakan perkiraan yang disarankan oleh (Breslow, 1984), yaitu dengan menyamakan *Pearson Chi-Square Statistic* dengan derajat bebas,

$$\sum_i \frac{\left(y_i - E(Y_i)\right)^2}{Var(Y_i)} = n - p$$

sehingga diperoleh,

$$\sum_i \frac{(y_i - \mu_i)^2}{\mu_i(1 + a\mu_i)} = n - p \tag{5.19}$$

di mana *n* menunjukkan jumlah *rating classes* dan *p* jumlah parameter regresi. Prosedur iterasi seperti yang disebutkan di atas juga dapat digunakan, kali ini menghasilkan MLE dari $\boldsymbol{\beta}$ dan estimasi moment dari a, $(\widehat{\boldsymbol{\beta}}, \tilde{a})$.

Dalam tulisan ini, ketika a diestimasi dengan MLE, model akan disebut sebagai binomial negatif (MLE). Demikian juga, ketika diestimasi dengan metode moment, model akan disebut sebagai binomial negatif I (moment).

Uji ketepatan (*goodness of fit*) model regresi binomial negatif

Uji ketepatan model regresi binomial negatif dilakukan dengan dua cara yaitu, *Pearson Chi-Square statistic* dan *deviance*.

Pearson Chi-Square pada model regresi binomial negatif

Karena data yang diobservasi itu berdistribusi binomial negatif, maka $Var(Y_i) = \mu_i(1 + a\mu_i)$ sehingga *Pearson Chi-Square Statistic* untuk regresi binomial negatif I yaitu:

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i(1 + a\mu_i)}. \tag{5.20}$$

Deviance pada model regresi binomial negatif

Uji kecocokan suatu model terhadap data adalah pertanyaan alami yang timbul pada semua model statistik. Salah satu cara untuk menilai kecocokan model adalah dengan membandingkannya dengan model penuh (*saturated model*). *Deviance* dinotasikan dengan *D*, didefinisikan sebagai ukuran jarak antara *saturated model* dengan *current model*:

$$D = 2(logL(\boldsymbol{y}; \boldsymbol{a}) - logL(\boldsymbol{\mu}; \boldsymbol{a})).$$

Bila model cocok, maka *current model* diharapkan dekat dengan (tapi tidak lebih besar dari) *saturated model.* Nilai *deviance* yang besar menunjukkan bahwa *current model* tidak bagus. Sehingga bentuk persamaan di atas menjadi,

$$D = 2\sum_{i=1}^{n}\left\{y_i\, log\left(\frac{y_i}{\mu_i}\right) - (y_i + a^{-1})\, log\left(\frac{1 + ay_i}{1 + a\mu_i}\right)\right\}. \tag{5.21}$$

Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder dari laporan Direktorat Lalu Lintas Polda Sumatera Barat tahun 2012. *Rating factor* dan *rating classes* yang dipakai dalam penelitian ini berdasarkan pada penelitian sebelumnya serta berdasarkan data yang tersedia di Direktorat tersebut. Pada penelitian ini peneliti membatasi masalah pada faktor penyebab utama kecelakaan yaitu faktor manusia.

 ISBN : 978 – 979 – 16353 – 9 – 4

Tabel 1 *Rating factors* dan *rating classes* untuk data penelitian

Rating Factors	*Rating Classes*
Kepemilikan SIM	• Memiliki SIM • Tidak memiliki SIM
Usia pengendara	• <17 tahun • 17-24 tahun • 25-35 tahun • >35 tahun
Jenis kelamin pengendara	• Laki-laki • Perempuan
Pendidikan pengendara	• < SD • SD • SMP • SMA • S1 • >S1
Kondisi fisik pengendara	• Lelah • Mabuk • Sakit • Sehat

Tabel 1 menunjukkan *rating factors* dan *rating classes* untuk jumlah kecelakaan. Dalam hal ini, terdapat $2 \times 4 \times 2 \times 6 \times 4 = 384$ perkalian *rating classes* yang diklasifikasikan menurut frekuensi kecelakaan. Data lengkap, yang berisi jumlah kecelakaan, *rating factors* dan *rating classes* ditampilkan pada lampiran.

Variabel dalam penelitian ini terdiri dari variabel respon dan prediktor. Adapun variabel yang digunakan pada penelitian ini adalah:

1. Variabel respon (*Dependent variable*)

Dalam penelitian ini yang menjadi variabel respon adalah data angka jumlah kecelakaan tahun 2012.

2. Variabel bebas (*independent variable*) atau variabel penjelas

Ada kalanya kita melakukan suatu regresi dimana variabel penjelas berupa data kualitatif. Jika data kualitatif tersebut memiliki *m* kategori, maka jumlah variabel dummy yang dicantumkan di dalam model adalah (*m* - 1) Sedangkan yang menjadi variabel penjelas utama pada penelitian ini adalah tidak memiliki SIM, 17-24 tahun, 25-35 tahun, >35 tahun, perempuan, SD, SMP, SMA, S1, >S1, mabuk, sakit dan sehat.

Proses Analisis Data

Untuk menguji signifikansi dan untuk menilai interaksi antara pasangan *rating factors* digunakan analisis *Deviance*. Berikut ini akan disajikan hasil analisis *Deviance* model regresi Poisson yang memuat *rating factors* utama dan semua interaksinya.

Dari simulasi terlihat bahwa terdapat pasangan rating factor yang tidak signifikan. Sehingga harus dilakukan kembali analisis *Deviance* dengan membuang pasangan rating factor yang tidak signifikan, untuk selanjutnya penulis menitikberatkan pada interaksi antara pendidikan pengemudi dan kodisi fisik pengemudi yang merupakan salah satu pasangan rating factor yang signifikan. Berikut adalah ini adalah analisis *Deviance* model regresi Poisson beserta data kombinasi antara pendidikan pengemudi dan kondisi fisik pengemudi.

ISBN : 978 – 979 – 16353 – 9 – 4

Tabel 2 Analisis *Deviance* model regresi Poisson untuk *rating factor* utama dan pasangan yang sudah signifikan.

	Df	*Deviance*	P(>|Chi|)	
NULL				
Kepemilikan_SIM	1	358.03	< 2.2e-16	***
Usia	3	20.71	0.000121	***
Pendidikan	5	22.55	0.0004117	***
Kondisi_Fisik	3	23.23	3.62E-05	***
Pendidikan:Kondisi_Fisik	15	64.95	3.49E-08	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Berdasarkan analisis *Deviance,* model terbaik adalah model yang semua faktor signifikan. Banyak *rating factor* sekarang berkurang dari lima menjadi empat.

Dari analisis *Deviance* antara *rating classes* diperoleh bahwa ternyata tidak semua *rating clases* signifikan, untuk itu peneliti harus melanjutkan analisis deviansi sehingga semua *rating clases* singnifikan. Seperti terlihat pada Tabel. 3.

Tabel. 3 Analisis *Deviance* model regresi Poisson untuk tiap *rating classes* yang telah signifikan

	Df	*Deviance*	P(>|Chi|)	
NULL				
A2	1	358.03	< 2.2e-16	***
D1_E2	1	11.07	0.0008767	***
D3_E1	1	11.73	0.0006137	***
D5_E3	1	14.16	0.000168	***
D6_E1	1	26.16	3.14E-07	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

Dari hasil analisis deviansi model regresi Poisson untuk tiap *rating classes* hanya diperoleh satu *rating classes* yang signifikan yaitu perihal tidak memiliki SIM dan empat pasangan *rating classes* yaitu kombinasi antara pendidikan pengemudi dan kondisi fisik pengemudi. Hasil estimasi parameter untuk *rating classes* yang signifikan, dapat dilihat pada Tabel. 4.

Tabel 4 Estimasi parameter untuk model Poisson *rating classes* yang signifikan

Nilai estimasi beta tiap dummy					
No.	**Beta**	**Nilai_beta**	**Standar.error**	**Varians**	**Pval**
1	Intercept	1.6	0.033	0.001	0
2	Tidak_memiliki SIM	0.72	0.039	0.002	0
3	<SD_Lelah	0.27	0.082	0.007	0.00117
4	SMP_Mabuk	-0.36	0.11	0.012	0.00108
5	S1_Sakit	0.3	0.081	0.007	0.00024
6	>S1_Mabuk	-0.57	0.122	0.015	0
	df	378			
	Pearson's X^2	713.814			
	Deviance	389.6122			
	log L	-1096.481			

 ISBN : 978 – 979 – 16353 – 9 – 4

AIC	2204.961
BIC	2228.665

Nilai *p-value* untuk semua parameter kecil dari 0.005, nilai ini mengidentifikasikan bahwa estimasi parameter sudah signifikan. Dari TTabel 4 terlihat bahwa terjadi overdispersi pada data karena nilai Pearson *Chi-square* dan *Deviance* dibagi dengan derajat bebas nilainya lebih besar dari 1. Untuk mengatasi masalah overdispersi pada data cacah ini dapat diatasi dengan memodel-ulangkan dengan regresi binomial negatif.

Tabel 5 Estimasi parameter untuk model Binomial Negatif *rating classes*

Nilai a = 12.94691					
Nilai estimasi beta tiap dummy					
No.	**Beta**	**nilai_beta**	**standar.error**	**varians**	**Pval**
1	Intercept	1.62	0.27	0.075	0
2	Tidak_memiliki SIM	0.7	0.37	0.136	0.05
3	<SD_Lelah	0.2	0.93	0.858	0.832
4	SMP_Mabuk	-0.27	0.93	0.862	0.767
5	S1_Sakit	0.23	0.93	0.857	0.802
6	>S1_Mabuk	-0.41	0.93	0.864	0.655
	df	377			
	Pearson's X^2	6.209498			
	Deviance	8.613413			
	log L	-1716.964			
	AIC	3445.927			
	BIC	3469.631			

Setelah data yang sama seperti data yang digunakan untuk Tabel.4 dimodel ulangkan dengan model Binomial Negatif, terlihat bahwa permasalahan overdispersi yang terjadi pada model Poison dapat teratasi. Hal ini dapat dilihat pada Tabel.5, dimana nilai Pearson *Chi-square* dan *Deviance* dibagi dengan derajat bebas nilainya kecil dari 1. Tetapi sebagian besar parameter pada data yang dimodel-ulangkan ini menjadi tidak signifikan.

D. SIMPULAN DAN SARAN

Dari hasil simulasi data kecelakaan yang terjadi di Sumatera Barat selama tahun 2012 dapat disimpulkan bahwa:

1. terdapat interaksi antara pendidikan pengemudi dan kodisi fisik pengemudi yang merupakan salah satu pasangan rating factor yang signifikan.
2. berdasarkan analisis *Deviance* antara *rating factor* ternyata jumlah *rating factor* sekarang berkurang dari lima menjadi empat.
3. Berdasarkan analisis *Deviance* antara *rating classes* diperoleh bahwa ternyata tidak semua *rating clases* signifikan, untuk itu harus dilanjutkan analisis deviansi sehingga semua *rating clases* singnifikan.
4. Berdasarkan hasil analisis deviansi model regresi Poisson untuk tiap *rating classes* hanya diperoleh satu *rating classes* yang signifikan yaitu perihal tidak memiliki SIM dan empat pasangan *rating classes* yaitu kombinasi antara pendidikan pengemudi dan kondisi fisik pengemudi.

 ISBN : 978 – 979 – 16353 – 9 – 4

5. Setelah dilakukan pemodelan ulang dengan model Binomial Negatif, terlihat bahwa permasalahan overdispersi yang terjadi pada model Poison dapat teratasi. Hal ini dapat dilihat pada nilai Pearson *Chi-square* dan *Deviance* dibagi dengan derajat bebas nilainya kecil dari 1. Tetapi sebagian besar parameter pada data yang dimodel-ulangkan ini menjadi tidak signifikan

Penelitian ini hanya melibatkan satu faktor utama yaitu manusia. Pada hal masih ada factor penyebab kecelakaan yang lain seperti: kondisi kendaraan, faktor cuaca, dan faktor jalan dengan berbagai *rating factor* dan *class factor*. Penelitian selanjutnya terbuka untuk faktor-faktor yang belum diungkap tersebut.

E. DAFTAR PUSTAKA

Bailey, Robert A., and Leroy J. Simon. 1960 "Two Studies in Automobile Insurance Ratemaking." ASTIN Bulletin, hal. 192-217.

Bain, Lee J., and Max Engelhardt. 1991. Introduction to Probability and Matematical Statistic. California: Duxbury Press.

Breslow, N. E. 1984. "Extra-Poisson Variation in Log-Linear Models." Journal of the Royal Statistical Society, Blackwell Publishing for the Royal Statistical Society.

Cameron, A. Colin, dan Pravin K. Trivedi. 1986. “Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests.” Journal of Applied Econometrics, hal. 29-53.

Cox, D. R. 1983. “Some Remarks on Overdispersion.” Biometrika, hal. 269-274.

Harrington, Scott E. 1986. "Estimation and Testing for Functional Form in Pure Premium Regression Models." ASTIN Bulletin, hal. 31-43.

Ismail, Noriszura, and Abdul Aziz Jemain. 2005. "Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation." Casualty Actuarial Society Forum, hal. 367-394.

Ismail, Noriszura, and Abdul Aziz Jemain. 2007. "Handling Overdispersion with Negative binomial and Generalized Poisson Regression Models." Casualty Actuarial Society Forum, hal. 103-158.

Jong, Piet de, and Gillian Z. Heller. Generalized Linear Models for Insurance Data. New York: Cambridge University Press, 2008.

Lawless, Jerald F. 1987. “Negative binomial and Mixed Poisson Regression.” The Canadian Journal of Statistics, hal. 209-225.

McCullagh, P., dan J.A. Nelder. 1989. Generalized Linear Models. 2nd Edition. London: Chapman and Hall.

Nelder, J. A., dan Y. Lee. 1992. “Likelihood, Quasi-Likelihood and Pseudolikelihood: Some Comparisons.” urnal of the Royal Statistical Society, hal. 273-284.

Renshaw, Arthur E. 1994. “Modelling the Claims Process in the Presence of Covariates.” ASTIN Bullletin, hal. 265-285.

Sari, Devni Prima. 2010. Penanganan Overdispersi dengan Model Regresi Binomial Negatif I Pada Faktor-Faktor yang Mempengaruhi Angka Kematian yang Disebabkan Oleh Kanker Paru-Paru. Prosiding Seminar Nasional Matematika UNS, hal. 445-452.

Sari, Devni Prima. 2010. Penanganan Overdispersi dengan Model Regresi Binomial Negatif I dan Binomial Negatif II. Yogyakarta: Universitas Gadjah Mada.

Sari, Devni Prima. 2011. Kajian Overdispersi Pada Regresi Poisson dengan Menggunakan Regresi Poisson Tergeneralisasi I. Prosiding Seminar Nasional Matematika Padang: UNAND, hal. 122-128.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." The Annals of Statistics, hal. 461-464.

UU RI No.22 Tahun 2009. Undang-Undang No. 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan. Diakses melalui www.polri.go.id tanggal 2 Januari 2013.

Winkelmann, Rainer. 2008. Econometric Analysis of Count Data. New York: Springer